DOCUMENT RESUME

ED 333 031 TM 016 532

AUTHOR Tallmadge, G. Kasten
TITLE Randomized Field Experiments in Education.
PUB DATE Apr 91
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attrition (Research Studies); Control Groups; Dropout Programs; *Educational Experiments; Educational Research; Experimental Groups; *Field Studies; Matched Groups; *Program Evaluation; Research Design; Research Methodology; *Research Problems; *Sample Size; Sampling
IDENTIFIERS *Randomized Experiments

ABSTRACT

Problems with conducting randomized field experiments in education are explored. Focus is on problems encountered while evaluating a group of dropout prevention projects. Project planners were asked to manipulate the subject eligibility criteria until they identified as eligible three to four times as many students as they could serve. They were instructed to recruit all of them, allowing for the possibility that between 25% and 50% of the those eligible would not apply. A lottery would be used to identify students to be served. Students not selected for the intervention might display "resentful demoralization," acting as though some negative intervention had occurred. Techniques to alleviate this demoralization are discussed. Other problems of attrition in subject numbers included non-random group reassignment by a project director, and failure to track students who left the school or the area. It is concluded that implementing randomized control group designs is difficult and that a perfect experiment of this type can not be implemented. However, with care, the approach can be managed. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333031

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. KASTEN TALLMADGE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Randomized Field Experiments in Education

G. Kasten Tallmadge
American Institutes for Research

Paper presented at the annual meeting of the American Educational Research Association. Chicago, April 7, 1991.

TM016532

BEST COPY AVAILABLE

I attempted to do a randomized field experiment to evaluate an alternative high school back in 1979. Even though I remembered Campbell and Stanley's (1966) advice that the proper thing to do was to make comparisons between the group of students assigned to the treatment group (whether or not they received any treatment) and the group of students assigned to the control group, that is not what I did. Eighty-two percent of the treatment group students either failed to show up or dropped out--and we absolutely could not find most of them with the resources we had. Almost the only members of the treatment group we could find were those that remained in the school. We were able to find a few members of the control group--but most of them were unwilling to respond to our data collection instruments even though we offered a $10 incentive. We ended up making what comparisons we could between the members of the two groups that we could find. Of course, the groups of students we could find were no longer randomly equivalent, and we had to treat the evaluation as a non-equivalent comparison group design. It is not clear that the evaluation derived any benefit whatsoever from the initial random assignment.

At the time of this bad experience, Cook and Campbell (1979) hadn't yet invented the resentful demoralization threat to the internal validity of experiments--or at least the book in which they first described it had not yet been released. But I fell victim to resentful demoralization anyway. One nice California day I received a telephone call from a New York lawyer who had been retained by a young woman I had randomly assigned to the control group. Obviously, she felt strongly that this was unfair treatment. The case never did go to court, but that possibility, I must admit, disturbed me. Partially for that reason, I subsequently advised the Ford, Rockefeller, and Clark foundations not to undertake random assignment evaluations of their programs. I also really believed that resentful demoralization was at least as serious a threat to the internal validity of randomized experiments as the much more widely known threats of history, maturation, selection, and all of their interactions are threats to the validity of quasi-experiments.

I am not sure how I feel about these matters today, but I am doing some randomized field experiments in evaluating dropout prevention programs. And I am running into implementation

problems. Some of them may be worth sharing--at least they are problems that Bob Boruch (Boruch & Cordray, 1980) never warned us about when he was so strongly advocating the use of the randomized control group design.

The projects we are evaluating were told when they applied for their grants that they would have to cooperate with the federal evaluator and implement either a randomized control group design or a matched comparison group design. Most of them opted for the matched comparison group design--but the feds really wanted to give the randomized control group design a try. So I found myself pushing the design and attempting to get people to agree to implement it.

In several instances I thought I had succeeded, but when I described what was involved, I lost many of my would-be converts. Most of them told me that they were able to serve all of the eligible applicants and wished to do so. But, as I later learned, eligibility is a matter of degree. It is problematic to assess.

The way people select students to participate in their dropout prevention projects usually proceeds in two stages as follows. A group of at-risk students is first identified from teacher or counselor referrals, from grade and attendance records, or from other indicators such as being overage for their grade placement. Once this pool of "eligible" students has been identified, recruitment (stage 2) begins. Usually during the course of face-to-face interviews, students are informed about the project and the services they could receive through it. Then they are asked to apply.

Random assignment, properly done, must await the completion of both of these two stages. Otherwise, the treatment group would consist entirely of volunteers while the control group would have a smaller proportion of students who would have volunteered. Unfortunately, once program personnel have gone through the recruiting process, they feel it is unfair to deny any of the recruited students admission to the program. They are quite willing to allow students to be randomly assigned after their eligibility has been determined, but not after they have been recruited. I am not going to speculate as to why this is so--but I have seen it often enough and in sites that are diverse enough so that I believe in it. Of course there is always the possibility

that the recruiters recruit most vigorously those students whom they think will do well in the program, thus presumably enhancing their chances of looking good in the evaluation.

Recruiting, at least in the larger school districts, is often done in the spring of the year for the following fall. Some randomly assigned members of the treatment and control groups inevitably fail to show up the following fall, when the program begins. Also inevitably, schedule conflicts are uncovered in the fall that were not anticipated the previous spring. Other members of the treatment group are thus lost to the experiment and one cannot assume that all of these losses occur randomly. The original, randomly equivalent groups can quickly be transformed into non-equivalent groups through this type of attrition. And the worst part about it all is that I, at least, can think of no strategy for preventing or controlling these potential biases. Of course, if the random assignment could be held off until spring when the pool of students could be limited to eligible applicants with no schedule conflicts, the problem would be solved. But school people tell me that option is impossible.

But let's get back to matters over which we may have some control (which, we will assume, does not include *when* the random assignment is done). My experience suggests that the best starting place is to convince project personnel that they must identify and recruit more students than they can serve. Many recruiters have had enough experience with needs assessment to know that their attempts to predict which students are the most at risk are less than perfectly valid. They are usually willing to admit that they miss some at-risk students and that they overestimate how much at risk others are. At that point they begin to soften up and become more willing to consider the possibility of letting chance play a role in the selection process.

I ask project staff to manipulate the eligibility criteria until they identify as eligible three to four times as many students as they can serve and to recruit all of them. Somewhere between one quarter and one half of the "eligible" students will decide not to apply--and recruiters are familiar with this source of attrition. So we are left with somewhat more than twice as many eligible applicants as can be served. It is at this point that we introduce a lottery and use it to identify the students that will be served.

Using a random selection process in this way seems to bother staff much more than it bothers students. Staff seem to feel that it is unfair not to serve the students they perceive as neediest. Students, on the other hand, seem to feel that pure chance is fairer than what they believe are often biased staff judgments.

Of course, students must be told that there will be a lottery when they are recruited. They should also be told that the reason for the lottery is that more students will be eligible than can be served. There will still be some resentful demoralization among students assigned to the control group, but it has not appeared to be too severe. One caution, however, should be communicated. I strongly advise against assigning students to a control group if there are going to be unfilled "slots" in the treatment group. We had this problem in the 1979 study I referred to earlier, and it is a really difficult one. Control group students who want to get into the project can get quite angry when assigned to the control group under these circumstances. They do not behave as they would if there were no treatment. They behave as if they had received some type of negative treatment. Assigning students to the control group under these circumstances should properly be regarded as hazardous--not only to the internal validity of the evaluation (resentfully demoralized students will not provide us with a valid no-treatment performance estimate), but to the evaluator's health as well.

During this past year, we have tried a strategy to minimize resentful demoralization--we have described the control group as a waiting list and promised control group students that they would be given the first opportunity to fill any openings that might occur in the treatment group. When such openings have occurred in the treatment group, we have then randomly selected members of the control group to fill them. Our original assumption, not surprisingly, was that the selected members of the control group would accept this opportunity--but some did not. We decided that, under such circumstances, we should not replace the replacement student--although that decision now seems somewhat questionable.

To preserve the integrity of the randomized control group design, all members of the originally assigned control group should have been regarded as members of that group for the duration of the study without regard for what subsequently happened to them. So we could have students who were assigned to the control group in

spring but who subsequently were selected to fill an opening in the treatment group when the treatment began the following fall. These students would have received the entire treatment--but would continue to be regarded as members of the control group. Conversely, we could have students who were assigned to the treatment group who remained in school but who elected not to participate in the treatment. Both of these classes of students would tend to lower the treatment-effect estimate compared to what it would have been had all students remained in the groups to which they were originally assigned. The post-treatment performance of the treatment group students who received no treatment should have been depressed compared to what it would have been had those students received the treatment, while the post-treatment performance of the control group students who actually received the full treatment would have been inflated.

Somehow, having members of the control group who had received the full treatment and members of the treatment group who had received no treatment seemed counterintuitive when we were forced to decide how to handle these cases. We opted to try to preserve what was left of the equivalence of the groups whose students remained where they had been assigned. So, when a control student was moved to the treatment group, or when a treatment group student decided not to participate in the treatment, we simply removed those students from the study sample.

This was one of those rare occasions where we could have had our cake and eaten it too. Had we defined our groups in terms of their original assignments, we could have conducted our analyses without any of the biases that would have been introduced by differential attrition. Then we could have gone through the database, culled the cases where control students received the treatment and where treatment students did not. Analyses of the purged database (and the culled data) would give us not only another estimate of the treatment effect but some insight into possible biases introduced by the reassignments that were made. The circumstance that mitigates the error of our decision is the fact that we would have lost some students no matter what we did to try to track them. If this attrition was systematically different for the two groups--and we believe that it probably would have been--then some bias would inevitably be present in the analytic findings.

We encountered several other problems which may be worth mentioning because they might have been avoidable. The director of one of our projects moved several students from the control group to the treatment group when openings occurred. These group reassignments, unfortunately, were not made randomly but because the students "really wanted to be in the treatment" or because the project director felt they "would do really well." More intensive training of this individual probably would have forestalled these reassignments.

The last problem to which I want to refer is the return of the presumed dropouts. We had a number of students who left our projects and were gone for a long enough period of time for us to believe that they had dropped out. No one in the schools knew what had happened to them, but sometime later they returned and wanted to be readmitted to the treatment group. In some instances, they were gone as long as a year, and we were reluctant to take them back into "the study's" treatment group. Of course, we had no objection to their receiving the treatment when they returned but, for the purposes of the study, we were inclined to consider them dropouts as of the time they left (in at least some cases, we had conducted exit interviews). In any case, because we had not tracked them to wherever they had gone, we had big gaps in our data (e.g., attendance, grade-point averages). We now feel that these students represent object lessons on the importance of tracking students who leave school (and the area). Attrition from the sample is a big enough problem without losing students who might have been retained. We could, of course, have quantified the amount of treatment they received, and built that into a regression equation relating amount of treatment received to benefits accrued.

To summarize, implementing randomized control group designs is difficult. Even with prior experience, we encountered problems that we had not anticipated. And, in retrospect, we made some mistakes. It appears to us now that it is probably impossible to implement a perfect true experiment in real-world settings. What we should strive to do, however, is to maximize our options. In other words, if one choice enables us to analyze the data in two different ways, we should choose that option rather than choosing an option that would cut us off from one or more alternative analytic options. Particularly if one of our somewhat flawed options would lead us to underestimate a treatment effect while another would lead us to overestimate it, there are obvious

benefits to doing both and thereby establishing boundaries within which the "correct" estimate must, presumably, lie.

REFERENCES

Boruch, R.F. & Cordray, D.S. (1980). *An appraisal of educational program evaluations: Federal, state, and local agencies.* Evanston, IL: Northwestern University. (U.S. Department of Education contract number 300-79-0467)

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research.* Chicago IL: Rand McNally

Cook, T.D. & Campbell, D.T. (eds.) (1979). *Quasi-experimentation: Design and analysis issues for field setings.* Skokie, IL: Rand McNally.